AD-A068 492 STANFORD UNIV CALIF SYSTEMS OPTIMIZATION LAB F/G 12/2

GLOBALLY CONVERGENT ALGORITHMS FOR CONVEX PROGRAMMING.(U)

FEB 79 E ROSENBERG N00014-75-C-0267

UNCLASSIFIED SOL-79-1 NL

1 OF 1

AD A068492

END

DATE FILMED

6 -- 79

DDC

Systems
Optimization
Laboratory

LEVEL II

DDC
RECEIVED
MAY 9 1979
C

AD A068492

DDC FILE COPY

This document has been approved for public release and sale; its distribution is unlimited.

Department of Operations Research
Stanford University
Stanford, CA 94305

79 05 09 052

(12)

SYSTEMS OPTIMIZATION LABORATORY
DEPARTMENT OF OPERATIONS RESEARCH
Stanford University
Stanford, California
94305

DDC
RECEIVED
MAY 9 1979
C

AD A068492

DDC FILE COPY

GLOBALLY CONVERGENT ALGORITHMS FOR CONVEX PROGRAMMING

by

Eric Rosenberg

See back page for 1473

TECHNICAL REPORT SOL 79-1
February 1979

Research and reproduction of this report were partially supported by the Department of Energy Contract DE-AS03-76SF00034; the Office of Naval Research Contract N00014-75-C-0267; and the National Science Foundation Grants MCS76-81259 A01 and MCS76-20019 A01.

Reproduction in whole or in part is permitted for any purposes of the United States Government. This document has been approved for public release and sale; its distribution is unlimited.

79 05 09 052

ACCESSION for
NTIS White Section
DDC Buff Section
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. SPECIAL
A

GLOBALLY CONVERGENT ALGORITHMS FOR CONVEX PROGRAMMING

by Eric Rosenberg

1. Introduction

Over the years, numerous algorithms have been proposed for minimizing a nonlinear objective function subject to nonlinear constraints. Many of these algorithms can be classified as primal approximation methods. These methods treat the given problem, hereafter referred to as the primal, by using the current estimate of a primal solution, possibly together with other information, such as estimates of the Lagrange multipliers, to form a constrained minimization subproblem which in some way approximates the primal. The procedure of solving a sequence of such approximating subproblems, and perhaps executing other tasks, we call recursive substitution. For example, with x_i as the current estimate of a primal solution, we might solve the quadratic subproblem obtained by linearizing each constraint and the objective function about x_i and adding to the objective function the term $(x-x_i)^t G_i(x-x_i)$, where G_i is a positive definite matrix that approximates the Hessian of the Lagrangian at a Karush-Kuhn-Tucker (K.K.T.) pair [5,6,22,29]. Various non-quadratic subproblems have also been proposed [13,16,24,27].

All of the algorithms proposed in these references are pure recursive substitution schemes, that is, schemes which set x_{i+1} equal to a solution of the approximating subproblem generated from x_i. Other methods require additional computation, such as a line search, to generate x_{i+1} [7,19,20,30]. Furthermore, the methods of [13,16,24, 27,29] are one-point methods, that is, methods that use only information

at the current point. Methods that use quasi-Newton updates [5,6,22] are not one-point methods, since the Hessian approximation depends on previous estimates of a K.K.T. pair.

It is reasonable to expect a recursive substitution scheme to be effective if each subproblem can be easily solved. Notice that in general a trade-off is inevitable: the easier a particular type of subproblem is to solve, the less it tends to resemble the primal, and consequently the more subproblems we expect to have to solve. The primal itself is of course a perfect approximation and presumably is difficult to solve, while the subproblem formed by linearizing the constraints and objective function can be easily solved by linear programming, but may be a poor approximation, especially if the functions defining the primal are highly nonlinear. In particular, approximating a geometric program by a linear program [3] can be disastrous, and often it is desirable to approximate a geometric program by another geometric program, whose constraints and objective function will in general be nonlinear functions [28]. In this case, each approximating geometric program has the advantage of a smaller degree of difficulty than the given problem. There is therefore a need to study general approximating subproblems.

In order for any algorithm to be used with confidence, it is necessary to determine under what conditions, if any, the algorithm generates a sequence of estimates that converges to a solution, and, if convergence can be established, it is important to determine the rate of convergence. Most algorithms popular today, and in particular

most pure recursive substitution schemes, exhibit local convergence. That is, for any starting primal solution estimate x_0 in some neighborhood of a primal stationary point z, the algorithm generates a sequence $\{x_i\}$ that converges to z. A local convergence proof generally requires strong differentiability assumptions and a good estimate of a vector of Lagrange multipliers at z. Rate of convergence results are necessarily local results, and in fact are usually established in the course of proving local convergence. In [25], local convergence and rate of convergence results are derived for methods utilizing arbitrary, possibly non-quadratic, approximating subproblems in a one-point recursive substitution scheme.

Few researchers have considered the question of global convergence. We will say that a nonlinear programming algorithm is globally convergent if, for any arbitrary starting primal solution estimate x_0, the algorithm generates a sequence $\{x_i\}$ that converges to a primal stationary point. A globally convergent algorithm is extremely desirable, for a locally convergent method might fail miserably if provided with a poor initial estimate, and a feasible direction method [33] requires an initial feasible point, which might be unavailable or difficult to compute.

Recently, a globally convergent algorithm employing quadratic subproblems has been proposed [7]. Under appropriate hypotheses, the solution of each quadratic subproblem is shown to generate a descent direction of an exact penalty function θ_ρ, where ρ is a fixed positive real number. That is, let x_i be the current estimate of a solution, let z_i solve the quadratic subproblem constructed from

x_i and some positive definite matrix, as described above, and let $d_i = z_i - x_i$. Then $D_{d_i}\theta_\rho(x_i) < 0$, where $D_d f(x)$ denotes the directional derivative of the function f at the point x in the direction d. The new estimate is then $x_{i+1} = x_i + \alpha_i d_i$, where

$$\theta_\rho(x_i + \alpha_i d_i) = \min_{0 \leq \alpha \leq \beta} \theta_\rho(x_i + \alpha d_i) \text{ and } 0 < \beta < +\infty.$$

We then solve the quadratic subproblem constructed from x_{i+1} and a new positive definite matrix, and continue in this fashion. For each sufficiently large ρ, this scheme is globally convergent. Moreover, by using Lagrange multiplier estimates and choosing each G_i properly, in some neighborhood of a K.K.T. pair (z,u) the line search can be omitted, and the pure recursive substitution scheme itself generates a sequence $\{(x_i,v_i)\}$ that converges to (z,u) [5,6,22].

In this paper we will generalize the results of [7] to prove global convergence for recursive substitution schemes utilizing arbitrary, possibly non-quadratic, approximating subproblems. Alternatively, our results can be viewed as the global version of the results of [25], without the restriction to one-point schemes. We will restrict our attention to solving a convex primal with convex subproblems so that we can employ the full power of convex analysis and thereby determine the minimum hypotheses needed to guarantee global convergence. In particular, the functions defining the primal and each subproblem need not be differentiable. Our results also prove global convergence of a new algorithm for geometric programming [28].

This paper is divided into six sections. In the next two sections we examine the connection between constrained optimization problems and exact penalty functions. In Section 4 we consider the directional derivative of the maximum of a finite collection of convex functions. We present the global convergence theorem in Section 5. Section 6 is devoted to concluding remarks.

2. Exact Penalty Functions: Part 1

Our goal is to solve *convex program* $\mathcal{C}$:

$$\text{minimize} \quad f_0(x)$$

$$\text{subject to} \quad f_k(x) \leq 0, \qquad k = 1,2,\ldots,p,$$

where, for each $k = 0,1,\ldots,p$, $f_k: R^m \to R$ is a convex function. The *solution set* of $\mathcal{C}$ is the set of all points that solve $\mathcal{C}$.

We associate with program $\mathcal{C}$ the *exact penalty function* θ_ρ, defined by

$$\theta_\rho(x) = f_0(x) + \rho \sum_{k=1}^{\rho} \max(0, f_k(x)) ,$$

where ρ is a positive real number. The *minimum set* of θ_ρ is the set of points that minimize θ_ρ. We call θ_ρ an exact penalty function because, if the functions $\{f_k\}$ are differentiable, then for each sufficiently large ρ the minimum set

of θ_ρ and the solution set of $\mathcal{C}$ coincide [8,14,21,31]. Notice that finite values of ρ suffice, in contrast to the classical exterior penalty function methods requiring that ρ approach infinity [4]. Exact penalty functions have been extensively studied since 1967; a good bibliography can be found in [8].

In this section and the next, we consider the relationship between a compact family of convex programs and their associated exact penalty functions. Inspired by [31], we impose no differentiability assumptions. The proof of Theorem 1 is also fashioned after [31].

We denote an infinite sequence in R^m by $\{x_i\}$. Where no confusion can arise, we also write $x = (x_1, x_2, \ldots, x_m)$. By $x \geq 0$ we mean $x_j \geq 0$, $j = 1,2,\ldots,m$. If $x, y \in R^m$, by $\langle x,y \rangle$ we mean $\sum_{j=1}^m x_j y_j$, by $\|x\|_\infty$ we mean $\max_{1 \leq j \leq m} |x_j|$, and by $\|\cdot\|$ we mean the Euclidean norm $\|\cdot\|_2$. If $X, Y \subset R^m$ and $a \in R$, by $X + Y$ we mean $\{x + y | x \in X \text{ and } y \in Y\}$, and by aX we mean $\{ax | x \in X\}$. Throughout this paper, all functions map all of R^m into R, for some m, and, unless otherwise noted, are finite valued. We say that the function $f: R^m \to R \cup \{+\infty\}$ is convex if its epigraph $\{(x,\alpha) | f(x) \leq \alpha\}$ is convex as a subset of R^{m+1}. We observe that a finite valued convex function is necessarily continuous (Corollary 10.1.1, [26]). We say that convex program $\mathcal{C}$ is superconsistent if for some x^0 we have $f_k(x^0) < 0$, $k = 1,2,\ldots,p$. The end of a proof will be denoted by $\otimes$.

We begin with a review of point-to-set maps [9,10,11,15,16, 17,18,32]. Let $S \subset R^n$ and $X \subset R^m$. A <u>point-to-set map</u> M sends the point s in S to the subset $M(s)$ of X. If $s \in S$, the map M is said to be <u>closed at</u> s if $\{s_i\} \subset S$, $s_i \to s$, $x_i \in M(s_i)$, and $x_i \to x$ imply $x \in M(s)$. The map M is said to be <u>uniformly bounded near</u> s if there is an open neighborhood N of s such that the set $\cup_{y \in N \cap S} M(y)$ is bounded, and M is said to be <u>nonempty near</u> s if there is an open neighborhood N of s such that $M(y)$ is nonempty whenever $y \in N \cap S$. If $T \subset S$, then M is said to be closed on T, uniformly bounded near T, or nonempty near T if for each s in T the map M is closed at s, uniformly bounded near s, or nonempty near s, respectively.

We shall treat the subset S of R^n as a perturbation space. For each fixed s in S, we consider program $\bar{C}(s)$:

$$\begin{aligned} &\underset{x}{\text{minimize}} \quad \bar{f}_0(x,s) \\ &\text{subject to} \quad \bar{f}_k(x,s) \leq 0, \qquad k = 1,2,\ldots,p, \end{aligned}$$

where

$$\bar{f}_k : R^m \times R^n \to R \,, \qquad k = 0,1,\ldots,p \,.$$

We define the following maps:

$$\Phi(s) = \{x \mid \bar{f}_k(x,s) \leq 0,\ k = 1,2,\ldots,p\} \,,$$

Φ is a feasible region map;

$$\omega(s) = \inf\{\bar{f}_0(x,s) \mid x \in \Phi(s)\} ,$$

ω is an optimal objective value function;

$$\Omega(s) = \{x \in \Phi(s) \mid \bar{f}_0(x,s) = \omega(s)\} ,$$

Ω is a solution set map;

$$L(u,s) = \inf\{\bar{f}_0(x,s) + \sum_{k=1}^{p} u_k \bar{f}_k(x,s) \mid x \in R^m\},$$

L is an infimal Lagrangian function;

$$U(s) = \{u \geq 0 \mid L(u,s) = \sup_{v \geq 0} L(v,s)\} ,$$

U is a Lagrange multiplier (optimal dual variable) map.

LEMMA 1. Let $\tilde{s} \in R^n$, and suppose that $\omega(\tilde{s})$ is finite, the functions $\bar{f}_0, \bar{f}_1, \ldots, \bar{f}_p$ are convex in x for each fixed s and jointly continuous in x and s, the set $\Omega(\tilde{s})$ is nonempty and bounded, and there is a point x^0 such that $\bar{f}_k(x^0,\tilde{s}) < 0$, $k = 1,2,\ldots,p$. Then Ω and U are closed at $\tilde{s}$ and nonempty and uniformly bounded near $\tilde{s}$. Moreover, ω is continuous at $\tilde{s}$.

Proof. See Lemmas 1 and 2, Hogan [10]. ⊗

LEMMA 2. Let $M: R^n \to R^m$ be a point-to-set map and let $f: R^m \to R$ be a function. Let $v(s) = \sup\{f(x) | x \in M(s)\}$. If M is closed at $\tilde{s}$ and uniformly bounded near $\tilde{s}$, and if f is continuous on R^m, then v is upper semicontinuous at $\tilde{s}$.

Proof. See Theorem 5, Hogan [9]. $\otimes$

For each s in S, we associate with program $\bar{C}(s)$ the exact penalty function $\bar{\theta}_\rho(\cdot,s)$, defined on R^m by

$$\bar{\theta}_\rho(x,s) = \bar{f}_0(x,s) + \rho \sum_{k=1}^{p} \max(0,\bar{f}_k(x,s)) .$$

THEOREM 1. Suppose program C is superconsistent. Let $\bar{f}_0, \bar{f}_1,\ldots,\bar{f}_p$ be functions jointly continuous on $R^m \times R^n$ such that for each fixed s and $k = 0,1,\ldots,p$ the function $\bar{f}_k(\cdot,s)$ is convex on R^m, and such that for each x, s, and $k = 1,2,\ldots,p$ we have $\bar{f}_k(x,s) \leq f_k(x)$. Let S be a nonempty and compact subset of R^n such that the solution set of $\bar{C}(s)$ is nonempty and bounded whenever $s \in S$. Then there is a positive number $\bar{\rho}_1$ such that, whenever $\rho \geq \bar{\rho}_1$ and $s \in S$, each minimum of the function $\bar{\theta}_\rho(\cdot,s)$ is also a solution of program $\bar{C}(s)$.

Proof. Since C is superconsistent, then for some x^0 we have $\bar{f}_k(x^0,s) \leq f_k(x^0) < 0$ for each s in S and $k = 1,2,\ldots,p$.

Hence, $\bar{\mathcal{C}}(s)$ is superconsistent for each s in S. Let

$$\alpha = \min_{x \in S} \max_{1 \leq k \leq p} \bar{f}_k(x^0, s) .$$

Then $\alpha \leq \max_{1 \leq k \leq p} f_k(x^0) < 0$.

By Lemma 1, the optimal value function ω is continuous on S. Therefore, $\beta = \min_{s \in S} \omega(s)$ for some finite number β. Let

$$\gamma = \max_{s \in S} \bar{f}_0(x^0, s)$$

and let

$$\bar{\rho}_1 = \frac{\beta - \gamma - 1}{\alpha} .$$

It is clear that $\bar{\rho}_1 > 0$. We claim that $\bar{\rho}_1$ is the desired threshold value. To see this, choose $\tilde{s}$ in S and choose $\rho \geq \bar{\rho}_1$. Suppose that the point v is infeasible for $\bar{\mathcal{C}}(\tilde{s})$. Since $\bar{f}_0(\cdot, \tilde{s})$ and $\bar{\theta}_\rho(\cdot, \tilde{s})$ agree at feasible points of $\bar{\mathcal{C}}(\tilde{s})$, to establish the theorem it suffices to find a point z which is feasible for $\bar{\mathcal{C}}(\tilde{s})$ such that $\bar{\theta}_\rho(z, \tilde{s}) < \bar{\theta}_\rho(v, \tilde{s})$, for then $\bar{\theta}_\rho(\cdot, \tilde{s})$ must attain its minimum in the feasible region of $\bar{\mathcal{C}}(\tilde{s})$.

By definition of x^0 and v, there is a point x^B on the line segment joining x^0 and v such that x^B is on the boundary of the feasible region of $\bar{\mathcal{C}}(\tilde{s})$. Let $K = \{k \in \{1, 2, \ldots, p\} \mid \bar{f}_k(x^B, \tilde{s}) = 0\}$, and let the auxiliary function φ be defined on R^m by $\varphi(x) = \bar{f}_0(x, \tilde{s}) + \rho \sum_{k \in K} \bar{f}_k(x, \tilde{s})$. It follows that $\varphi(x^B) = \bar{\theta}_\rho(x^B, \tilde{s})$ $= \bar{f}_0(x^B, \tilde{s})$.

Since $\bar{f}_k(v, \tilde{s}) \geq 0$ whenever $k \in K$, we have

$$\sum_{k\in K} \bar{f}_k(v,\tilde{s}) = \sum_{k\in K} \max(0,\bar{f}_k(v,\tilde{s}) \leq \sum_{k=1}^{p} \max(0,\bar{f}_k(v,\tilde{s}).$$

Therefore,

$$\bar{f}_0(v,\tilde{s}) + \rho \sum_{k\in K} \bar{f}_k(v,\tilde{s}) \leq \bar{f}_0(v,\tilde{s}) + \rho \sum_{k=1}^{p} \max(0,\bar{f}_k(v,\tilde{s})),$$

or equivalently, $\varphi(v) \leq \bar{\theta}_\rho(v,\tilde{s})$.

To prove the theorem, it now suffices to show that $\varphi(x^B) < \varphi(v)$. We first show that $\varphi(x^0) < \varphi(x^B)$. Since $\alpha < 0$ and $\rho \geq \bar{\rho}_1$, it follows that $\rho = (\beta-\gamma-1-\epsilon)/\alpha$ for some nonnegative number ϵ. We have

$$\varphi(x^0) = \bar{f}_0(x^0,\tilde{s}) + \left(\frac{\beta-\gamma-1-\epsilon}{\alpha}\right) \sum_{k\in K} \bar{f}_k(x^0,\tilde{s})$$

$$\leq \bar{f}_0(x^0,\tilde{s}) + \left(\frac{\beta-\gamma-1-\epsilon}{\alpha}\right) \max_{1\leq k\leq p} \bar{f}_k(x^0,\tilde{s})$$

(since $\bar{f}_k(x^0,\tilde{s}) < 0$ for $k = 1,2,\ldots,p$)

$$\leq \bar{f}_0(x^0,\tilde{s}) + \beta - \gamma - 1 - \epsilon$$

(since $0 < \alpha^{-1} \max_{1\leq k\leq p} \bar{f}_k(x^0,\tilde{s}) \leq 1$ and $\beta - \gamma - 1 - \epsilon < 0$)

$$\leq \min_{s\in S} \omega(s) - 1 + \bar{f}_0(x^0,\tilde{s}) - \max_{s\in S} \bar{f}_0(x^0,s)$$

$$\leq \min_{s\in S} \omega(s) - 1 < \bar{f}_0(x^B,\tilde{s}) = \varphi(x^B),$$

that is, $\varphi(x^0) < \varphi(x^B)$.

Since φ is convex and $x^B = tx^0 + (1-t)v$ for some t in $(0,1)$, we have

$$\varphi(x^B) \leq t\varphi(x^0) + (1-t)\,\varphi(v) < t\varphi(x^B) + (1-t)\,\varphi(v)\ ,$$

or equivalently, $\varphi(x^B) < \varphi(v)$. Since we have shown that $\varphi(v) \leq \bar{\theta}_\rho(v,\tilde{s})$ and since $\varphi(x^B) = \bar{\theta}_\rho\,(x^B,\tilde{s})$, it follows that $\bar{\theta}_\rho(x^B,\tilde{s}) < \bar{\theta}_\rho(v,\tilde{s})$, which proves the theorem. $\otimes$

COROLLARY 1.1. Suppose that program $\mathcal{C}$ is superconsistent and has a nonempty solution set. Then there is a positive number ρ_1 such that, whenever $\rho \geq \rho_1$, each minimum of the function θ_ρ is also a solution of program $\mathcal{C}$.

Proof. The result follows from Theorem 1 by deleting all references to the variable s and the set S and replacing each $\bar{f}_k(x,s)$ with $f_k(x)$ for $k = 0,1,\ldots,p$. Notice that the result holds even if the solution set of $\mathcal{C}$ is unbounded. (This corollary appears in [31].) $\otimes$

3. Exact Penalty Functions: Part 2.

To prove the converse of Theorem 1, we will require several results from convex analysis [26]. Let f be a convex function. The vector x^* is said to be a subgradient of f at the point x if $f(y) \geq f(x) + \langle x^*,y-x\rangle$ for every y. The set of all subgradients of

f at x is called the <u>subdifferential</u> of f at x, and is denoted by $\partial f(x)$. For each x, $\partial f(x)$ is a nonempty and compact convex set (Theorem 23.4, [26]). Moreover, f is differentiable at x if and only if $\partial f(x) = \{\nabla f(x)\}$. Clearly, $\partial f(\alpha x) = \alpha\, \partial f(x)$ for each x and each positive number α.

<u>LEMMA 3</u>. Let $f_1, f_2, \ldots, f_n$ be convex functions and let $f = f_1 + f_2 + \cdots + f_n$. Then for each x we have

$$\partial f(x) = \partial f_1(x) + \partial f_2(x) + \cdots + \partial f_n(x) .$$

<u>Proof</u>. See Theorem 23.8, Rockafellar [26]. ⊗

We say that the function $f: R^m \to R \cup \{+\infty\}$ is <u>proper</u> if f is convex and if $f(x) < +\infty$ for at least one x. If $f: R^m \to R \cup \{+\infty\}$ is a convex function, we define the <u>closure</u> of f to be that function whose epigraph is the closure in R^{m+1} of the epigraph of f. It follows that a proper convex function is closed if and only if it is lower semicontinuous.

Let $f: R^m \to R \cup \{+\infty\}$ be a convex function. The <u>conjugate function</u> f^* is defined on R^m by $f^*(x^*) = \sup\{\langle x^*, x\rangle - f(x) \mid x \in R^m\}$. The conjugate f^* is a closed convex function, proper if and only if f is proper. If f is a closed proper convex function, then the conjugate of f^* is f, that is, $(f^*)^* = f$. Therefore, the conjugacy operation $f \to f^*$ induces a one-to-one symmetric correspondence in the class of all closed proper convex functions on R^m. Since f^*

may not be finite valued, the subdifferential $\partial f^*(x^*)$ may be empty for some x^*. However, if f is a closed proper convex function, then $x \in \partial f^*(x^*)$ if and only if $x^* \in \partial f(x)$.

If X is a convex set in R^m, the <u>indicator</u> <u>function</u> of X, denoted by $\delta(\cdot|X)$, is defined on R^m by

$$\delta(x|X) = \begin{cases} 0 & \text{if } x \in X \\ +\infty & \text{otherwise.} \end{cases}$$

There is an obvious one-to-one correspondence between a convex set and its indicator function, namely, $\delta(x|X_1) = \delta(x|X_2)$ for every x if and only if $X_1 = X_2$. The conjugate transform of $\delta(\cdot|X)$ is called the support function of X. We have $\delta^*(x^*|X) = \sup\{\langle x^*,x\rangle - \delta(x|X)\,|\,x \in R^m\}$ $= \sup\{\langle x^*,x\rangle\,|\,x \in X\}$. If X is also closed, then $\delta(\cdot|X)$ and $\delta^*(\cdot|X)$ are conjugate to each other (Theorem 13.2, [26]). Therefore, if X_1 and X_2 are closed convex sets, we have $\delta^*(x^*|X_1) = \delta^*(x^*|X_2)$ for every x^* if and only if $X_1 = X_2$.

Let f be a (finite valued) convex function. It can be shown (Theorem 23.4, [26]) that, for each x and d and each sequence $\{\alpha_i\} \subset R$ with $0 < \alpha_{i+1} \leq \alpha_i$ and $\lim_{i\to\infty} \alpha_i = 0$,

$$\lim_{i \to \infty} \frac{f(x + \alpha_i d) - f(x)}{\alpha_i}$$

exists and is finite. We call this limit the <u>directional</u> <u>derivative</u>

of f in the direction d at the point x, and denote it by $D_d f(x)$. Moreover, for each fixed x and d,

$$g(\alpha) = \frac{f(x + \alpha d) - f(x)}{\alpha}$$

is nondecreasing on $\{\alpha \in R | \alpha > 0\}$; also, for each $\alpha > 0$ we have $D_{\alpha d} f(x) = \alpha D_d f(x)$. We say that the direction d is a <u>descent direction</u> of f at the point x if $D_d f(x) < 0$, in which case the continuity of f implies $f(x + \alpha d) < f(x)$ for all sufficiently small positive α.

<u>LEMMA 4</u>. Let f be a convex function. Then for each x and d we have $D_d f(x) = \max\{\langle x^*, d\rangle | x^* \in \partial f(x)\}$.

<u>Proof</u>. See Theorem 23.4, Rockafellar [26]. ⊗

Although a stronger version of our next theorem appears in [23], our result has a particularly simple proof and is adequate for our purposes.

<u>THEOREM 2</u>. Let f be a convex function and let $g = \max(0,f)$. Then $\partial g(x)$ is nonempty for every x and

i) $\partial g(x) = \{0\}$ if $f(x) < 0$

ii) $\partial g(x) \supset \{\alpha x^* | 0 \leq \alpha \leq 1 \text{ and } x^* \in \partial f(x)\}$ if $f(x) = 0$

iii) $\partial g(x) = \partial f(x)$ if $f(x) > 0$.

Proof. It follows from the above remarks that $\partial g(x)$ is nonempty, closed, convex, and bounded for all x.

i) Suppose $f(x) < 0$. Then for each z in some neighborhood of x we have $g(z) = 0$. Therefore, for each d we have

$$0 = D_d g(x) = \max\{\langle x^*, d\rangle \mid x^* \in \partial g(x)\} .$$

It follows that $\partial g(x) = \{0\}$, which proves i).

ii) Suppose $f(x) = 0$. Choose x^* in $\partial f(x)$. Then

$$f(y) \geq f(x) + \langle x^*, y-x\rangle = \langle x^*, y-x\rangle$$

for each y. If $\langle x^*, y-x\rangle \geq 0$, then $f(y) \geq 0$, so that

$$g(y) = f(y) \geq f(x) + \langle x^*, y-x) \geq g(x) + \langle \alpha x^*, y-x\rangle$$

for each α in $[0,1]$; hence $\alpha x^* \in \partial g(x)$. On the other hand, if $\langle x^*, y-x\rangle < 0$, then

$$\begin{aligned} g(y) &= \max(0, f(y)) \geq \max(0, \langle x^*, y-x\rangle) = 0 \\ &= \max(0, \langle \alpha x^*, y-x\rangle \geq g(x) + \langle \alpha x^*, y-x\rangle \quad \text{for each } \alpha \geq 0. \end{aligned}$$

Hence $\alpha x^* \in \partial g(x)$ whenever $\alpha \in [0,1]$, which proves ii).

iii) Suppose $f(x) > 0$. Then for each z in some neighborhood of x we have $g(z) = f(z)$. Therefore, for each d we have $D_d f(x) = D_d g(x)$. It follows from Lemma 4 and the above remarks that $\partial g(x) = \partial f(x)$, which proves iii). $\otimes$

A remarkable feature of convex programming is the existence of necessary and sufficient conditions for optimality, even in the absence of differentiability. Consider again convex program $\mathcal{C}$:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_k(x) \leq 0, \quad k = 1,2,\ldots, p. \end{array}$$

If $\mathcal{C}$ is actually an unconstrained minimization problem, we call the solution set the <u>minimum set</u>.

<u>LEMMA 5</u>. Let f be a convex function. Then the minimum set of f is $f^*(0)$; in particular, the infimum of f is attained if and only if $\partial f^*(0)$ is nonempty.

<u>Proof</u>. See Theorem 27.1, Rockafellar [26]. $\otimes$

We say that a vector u in R^p is a vector of <u>Lagrange multipliers</u> for $\mathcal{C}$ if $u \geq 0$ and if the infimum of the function $f_0 + u_1 f_1 + \cdots + u_p f_p$ is finite and equal to the optimal objective function value of $\mathcal{C}$. We define the <u>Lagrangian</u> function L on $R^m \times R^p$ by

$$L(x,v) = \begin{cases} f_0(x) + v_1 f_1(x) + \cdots + v_p f_p(x) & \text{if } v \geq 0 \\ -\infty & \text{otherwise.} \end{cases}$$

The pair (z,u) is said to be a <u>saddlepoint</u> of L (with respect to maximizing in v and minimizing in x) if for every x and v we have $L(z,v) \leq L(z,u) \leq L(x,u)$.

LEMMA 6. The point z solves $\mathcal{C}$ and u is a vector of Lagrange multipliers for $\mathcal{C}$ if and only if (z,u) is a saddlepoint of the Lagrangian L. This condition holds if and only if the following Karush-Kuhn-Tucker (K.K.T.) conditions hold:

i) $u_k \geq 0$ and $f_k(z) \leq 0$, $k = 1,2,\ldots,p$

ii) $u_k f_k(z) = 0$, $k = 1,2,\ldots,p$

iii) $0 \in \partial f_0(z) + u_1\, \partial f_1(x) + \cdots + u_p\, \partial f_p(z)$.

(If i), ii), iii) hold, we call (z,u) a K.K.T. pair.) Moreover, if $\mathcal{C}$ is superconsistent, then z solves $\mathcal{C}$ if and only if there is a vector u such that (z,u) is a saddlepoint of the Lagrangian L (or equivalently, if (z,u) is a K.K.T. pair), and the set of Lagrange multipliers is identical to the set of points maximizing (over all v) the function $\min_{x \in R^m} L(x,v)$.

Proof. See Theorems 28.2 and 28.3 and Corollaries 28.3.1 and 28.4.1, Rockafellar [26]. $\otimes$

We now prove the main result of this section, the converse of Theorem 1.

THEOREM 3. Under the hypotheses of Theorem 1, there is a nonnegative number $\bar{\rho}_2$ such that, whenever $\rho \geq \bar{\rho}_2$ and $s \in S$, each solution of program $\bar{\mathcal{C}}(s)$ is also a minimum of the function $\bar{\theta}_\rho(\cdot,s)$.

Proof. By Lemma 1, for each s in S the set of Lagrange multipliers $U(s)$ is nonempty and closed at s, and uniformly bounded near s. Therefore, by Lemma 2, for some nonnegative number $\bar{\rho}_2$ we have $\bar{\rho}_2 = \max_{s \in S}\{\|u\|_\infty \mid u \in U(s)\}$.

We claim that $\bar{\rho}_2$ is the desired constant. To see this, choose $\tilde{s}$ in S and $\rho \geq \bar{\rho}_2$. Let $\tilde{z}$ be a solution of $\mathcal{C}(\tilde{s})$ and let u belong to $U(\tilde{s})$. Let

$$K_- = \{k \in \{1,2,\ldots,p\} \mid \bar{f}_k(\tilde{z},\tilde{s}) < 0\} ,$$

$$K_0 = \{k \in \{1,2,\ldots,p\} \mid \bar{f}_k(\tilde{z},\tilde{s}) = 0\} ,$$

and

$$K_+ = \{k \in \{1,2,\ldots,p\} \mid \bar{f}_k(\tilde{z},\tilde{s}) > 0\} .$$

Then K_+ is empty and $u_k = 0$ for each k in K_-, by Lemma 6. Hence, by Lemma 6 again, we have

$$\begin{aligned} 0 &\in \partial\bar{f}_0(\tilde{z},\tilde{s}) + \sum_{k=1}^{p} u_k\, \partial\bar{f}_k(\tilde{z},\tilde{s}) = \partial\bar{f}_0(\tilde{z},\tilde{s}) + \sum_{k\in K_0} u_k\, \partial\bar{f}_k(\tilde{z},\tilde{s}) \\ &\subset \partial\bar{f}_0(\tilde{z},\tilde{s}) + \rho \sum_{k\in K_0} \{\alpha_k x_k^* \mid 0 \leq \alpha_k \leq 1 \text{ and } x_k^* \in \partial\bar{f}_k(\tilde{z},\tilde{s})\} \\ &\subset \bar{\theta}_\rho(\tilde{z},\tilde{s}) , \end{aligned}$$

where the final assertion follows from Theorem 2 and Lemma 3. Therefore, $0 \in \partial\bar{\theta}_\rho(\tilde{z},\tilde{s})$, which implies that $\tilde{z} \in \partial\bar{\theta}_\rho^*(0,\tilde{s})$. By Lemma 5, $\tilde{z}$ is a minimum of $\bar{\theta}_\rho(\cdot,\tilde{s})$, which proves the theorem. ⊗

COROLLARY 3.1. Under the hypotheses of Theorem 1, there is a positive number $\bar{\rho}_3$ such that, whenever $\rho \geq \bar{\rho}_3$ and $s \in S$, the solution set of $\bar{\mathcal{C}}(s)$ and the minimum set of $\bar{\theta}_\rho(\cdot,s)$ coincide.

Proof. In light of Theorems 1 and 3, it suffices to choose $\bar{\rho}_3 = \max(\bar{\rho}_1,\bar{\rho}_2)$. $\otimes$

COROLLARY 3.2. Suppose that program $\mathcal{C}$ is superconsistent and has a nonempty and bounded solution set. Then there is a positive number ρ_3 such that, whenever $\rho \geq \rho_3$, the solution set of $\mathcal{C}$ and the minimum set of θ_ρ coincide.

Proof. Half of this corollary follows directly from Corollary 1.1. The other set containment follows from Theorem 3 by deleting all references to the variable s and the set S, and replacing each $\bar{f}_k(x,s)$ with $f_k(x)$ for $k = 0,1,\ldots,p$. $\otimes$

4. Directional Derivatives

We next consider the directional derivative of the maximum of a finite collection of convex functions.

LEMMA 7. Let $f:R^m \to R$ be a convex function. Then for each x and each $\epsilon > 0$ there is a $\delta > 0$ such that $\partial f(y) \subset \partial f(x) + \epsilon B$ whenever $y \in x + \delta B$, where $B = \{z \in R^m \mid \|z\| \leq 1\}$, the unit ball in R^m.

Proof. See Corollary 24.5.1, Rockafellar [26]. $\otimes$

LEMMA 8. Let $f: R^m \to R$ be a convex function. Then the point-to-set map ∂f is closed and uniformly bounded on R^m.

Proof. Since $\partial f(x)$ is bounded for each x, it follows immediately from Lemma 7 that ∂f is uniformly bounded on R^m.

To show that ∂f is closed, let $x_i \to x$, let $x_i^* \in \partial f(x_i)$, and let $x_i^* \to x^*$. Then for every y we have $f(y) \geq f(x_i) + \langle x_i^*, y-x_i \rangle$ for every i. Since f is continuous, it follows that $f(y) \geq f(x) + \langle x^*, y-x \rangle$. Therefore, $x^* \in \partial f(x)$. $\otimes$

LEMMA 9. Let $\{\beta_i\}$ be an infinite sequence of real numbers such that $\beta_{i+1} \leq \beta_i$ for every i. Suppose some subsequence $\{\beta_j\}$ converges to some number β. Then the entire sequence $\{\beta_i\}$ converges to β.

Proof. Choose $\epsilon > 0$. Since $\lim_{j \to \infty} \beta_j = \beta$, for some N_1 we have $\beta_{N_1} \leq \beta + \epsilon$. Hence,

$$\limsup_{i \to \infty} \beta_i \leq \beta_{N_1} \leq \beta + \epsilon \; .$$

On the other hand, we must have $\beta \leq \beta_i$ for every i, which implies that $\beta \leq \liminf_{i \to \infty} \beta_i$. Thus for any $\epsilon > 0$ we have

$$\beta \leq \liminf_{i \to \infty} \beta_i \leq \limsup_{i \to \infty} \beta_i \leq \beta + \epsilon \; ,$$

which proves the result. $\otimes$

The proof of the following principal result of this section is fashioned after [2].

THEOREM 4. Let K be a finite set, and let the functions $\{f(\cdot,k) \mid k \in K\}$ be convex on R^m. Let g be defined on R^m by $g(x) = \max\{f(x,k) \mid k \in K\}$, and let $I(x) = \{k \in K \mid f(x,k) = g(x)\}$. Then for each x and d the directional derivative $D_d g(x)$ is finite, and $D_d g(x) = \max\{D_d f(x,k) \mid k \in I(x)\}$.

Proof. Choose x and d. Since g is convex, the directional derivative $D_d g(x)$ exists and is finite. Since $D_{\alpha d} g(x) = \alpha D_d g(x)$ whenever $\alpha > 0$, it suffices to prove the result for the case $\|d\| = 1$. Let $x_i = x + \alpha_i d$, where $0 < \alpha_{i+1} \leq \alpha_i$ and $\lim_{i \to \infty} \alpha_i = 0$, and $\|d\| = 1$. Then for each i we have $(x_i - x)/(\|x_i - x\|) = d$ (since $\|d\| = 1$, we can interpret the components of d as direction cosines). Choose k_i in $I(x_i)$ and choose k in $I(x)$.

For each i we have

$$\frac{g(x_i) - g(x)}{\alpha_i}$$

$$= \frac{f(x_i,k_i) - f(x_i,k)}{\alpha_i} + \frac{f(x_i,k) - f(x,k)}{\alpha_i}$$

$$\geq \frac{f(x_i,k) - f(x,k)}{\alpha_i} \qquad \text{(since } k_i \in I(x_i))$$

$$\geq \frac{\langle x^*, x_i - x\rangle}{\alpha_i} \qquad \text{for every } x^* \text{ in } \partial f(x,k) .$$

Since $(x_i - x)/\alpha_i = d$ for every i, we have

$$\frac{g(x_i) - g(x)}{\alpha_i} \geq \langle x^*, d\rangle \quad \text{for every } x^* \text{ in } f(x,k),$$

or equivalently,

$$\frac{g(x_i) - g(x)}{\alpha_i} \geq \sup\{\langle x^*, d\rangle \mid x^* \in \partial f(x,k)\} = D_d\, f(x,k) .$$

Since this holds for each i and each k in $I(x)$, it follows that

$$D_d g(x) = \lim_{i \to \infty} \frac{g(x_i) - g(x)}{\alpha_i} \geq \max\{D_d f(x,k) \mid k \in I(x)\} .$$

To prove the reverse inequality, we first observe that, since K is finite, for some subsequence x_j (where $j \to +\infty$) and some k^0 in K we have $k^0 \in I(x_j)$. Since $f(\cdot,k)$ and g are continuous functions for each k, it follows that

$$g(x) = \lim_{j \to \infty} g(x_j) = \lim_{j \to \infty} f(x_j,k^0) = f(x,k^0) .$$

Therefore, $k^0 \in I(x)$. For each j, we have,

$$\frac{g(x_j) - g(x)}{\alpha_j} = \frac{f(x_j,k^0) - f(x,k^0)}{\alpha_j} \leq \frac{\langle x_j^*, x_j - x\rangle}{\alpha_j} = \langle x_j^*, d\rangle$$

for every x_j^* in $\partial f(x_j,k^0)$. That is,

$$\frac{g(x_j) - g(x)}{\alpha_j} \leq \sup\{\langle x_j^*, d\rangle \mid x_j^* \in \partial f(x_j,k^0)\} = D_d f(x_j,k^0) .$$

Since $\partial f(\cdot,k^0)$ is closed at x and uniformly bounded near x, it follows from Lemma 2 that (for fixed d) the directional derivative $D_d f(x,k^0)$ is upper semicontinuous at x. Therefore, by Lemma 9, we have

$$D_d g(x) = \lim_{i\to\infty} \frac{g(x_i) - g(x)}{\alpha_i}$$

$$= \lim_{j\to\infty} \frac{g(x_j) - g(x)}{\alpha_j} \le \limsup_{j\to\infty} D_d f(x_j,k^0)$$

$$\le D_d f(x,k^0) \le \max\{D_d f(x,k) \mid k \in I(x)\},$$

where the final inequality follows as $k^0 \in I(x)$. Thus we have shown that

$$\max\{D_d f(x,k) \mid k \in I(x)\} \le D_d g(x) \le \max\{D_d f(x,k) \mid k \in I(x)\} ,$$

which proves the theorem. $\otimes$

5. The Global Convergence Theorem

In Sections 2 and 3, we considered the family $\{\bar{\mathcal{C}}(s) \mid s \in S\}$ of constrained minimization problems, where the perturbation space S is a compact subset of R^n. Suppose now that, for each s, $\bar{\mathcal{C}}(s)$ is easily constructed from $\mathcal{C}$, $\bar{\mathcal{C}}(s)$ resembles $\mathcal{C}$, and $\bar{\mathcal{C}}(s)$ is easier to solve than $\mathcal{C}$. We may then regard $\bar{\mathcal{C}}(s)$ as an approximating subproblem, and expect that its solution helps us to solve $\mathcal{C}$. Indeed, we will show that, under appropriate hypotheses, solving $\bar{\mathcal{C}}(s)$ generates a descent direction of θ_ρ, the exact penalty function for the primal $\mathcal{C}$.

In primal approximation methods, the perturbation s always supplies an estimate of a primal solution, and may also supply other information, such as an approximation of the Hessian of the Lagrangian. Accordingly, we will write $S = Y \times W$, where $Y \subset R^m$ and $W \subset R^q$ for some $q \geq 0$. If $s \in S$, we will write $s = (y,w)$, where $y \in Y$ and $w \in W$. By $q = 0$ we mean $W = \emptyset$, in which case we disregard W and w, so that $(y,w) \in Y \times W$ will mean $y \in Y$. For each $k = 0,1,\ldots,p$ we will write $\bar{f}_k(x,y,w)$, instead of $\bar{f}_k(x,s)$. We will also write $\bar{\mathcal{C}}(s)$ as $\bar{\mathcal{C}}(y,w)$:

$$\begin{aligned} &\underset{x}{\text{minimize}} \quad \bar{f}_0(x,y,w) \\ &\text{subject to} \quad \bar{f}_k(x,y,w) \leq 0 , \qquad k = 1,2,\ldots,p . \end{aligned}$$

In constructing $\bar{\mathcal{C}}(y,w)$, we will always set y equal to the current estimate of a primal solution, while w may be any arbitrary element of the compact, possibly empty, set W.

For instance, we can generate quadratic subproblems as follows [7]. Suppose the functions $\{f_k\}$ defining $\mathcal{C}$ are continuously differentiable. Let a and b be positive numbers, and let $\mathcal{G}$ be the collection of symmetric $m \times m$ matrices satisfying

$$a\|x\|^2 \leq \langle x,Gx \rangle \leq b\|x\|^2 \qquad \text{for every } x .$$

Let x_i be the current estimate of a primal solution and let $G_i \in \mathcal{G}$ be the current approximation of the Hessian of the Lagrangian. We form the quadratic subproblem $QP(x_i,G_i)$ by defining

$$\bar{f}_0(x,x_i,G_i) = f_0(x_i) + \langle \nabla f_0(x_i),\ x-x_i \rangle + \frac{1}{2} \langle x-x_i, G_i(x-x_i) \rangle$$

and

$$\bar{f}_k(x,x_i,G_i) = f_k(x_i) + \langle \nabla f_k(x_i), x-x_i \rangle , \quad k = 1,2,\dots,p .$$

Returning to the general case, the following theorem shows the crucial role of each approximating subproblem $\bar{\mathcal{C}}(y,w)$. If $\bar{f}_k: R^m \times R^m \times R^q \to R$, we denote by $\partial_1 \bar{f}(\cdot,y,w)$ the subdifferential map with respect to the first argument, so that $\partial_1 \bar{f}(x,y,w) \subset R^m$.

<u>THEOREM 5</u>. Suppose program $\mathcal{C}$ is superconsistent. Let $\bar{f}_0, \bar{f}_1, \dots, \bar{f}_p$ be functions jointly continuous on $R^m \times R^m \times R^q$ such that for each fixed y, w, and $k = 0,1,\dots,p$ the function $\bar{f}_k(\cdot,y,w)$ is convex and $\partial_1 \bar{f}_k(y,y,w) = \partial f_k(y)$, and such that for each x, y, w and $k = 1,2,\dots,p$ we have $\bar{f}_k(x,y,w) \le f_k(x)$ and $\bar{f}_k(y,y,w) = f_k(y)$. Let Y be a nonempty and compact subset of R^m and let W be a compact subset of R^q such that program $\bar{\mathcal{C}}(y,w)$ has the unique solution $z(y,w)$ whenever $(y,w) \in Y \times W$. Let $d(y,w) = z(y,w) - y$. Then there is a positive number $\bar{\rho}_3$ such that $D_{d(y,w)}\theta_\rho(y) < 0$ whenever $\rho \ge \bar{\rho}_3$, $(y,w) \in Y \times W$, and $d(y,w) \ne 0$.

<u>Proof</u>. We claim that the value $\bar{\rho}_3 = \max(\bar{\rho}_1, \bar{\rho}_2)$ specified in Corollary 3.1, with $S = Y \times W$, is a satisfactory choice. To see this, choose $(\tilde{y},\tilde{w})$ in $Y \times W$ and $\rho \ge \bar{\rho}_3$. Let $\tilde{z} = z(\tilde{y},\tilde{w})$, the unique solutions of $\bar{\mathcal{C}}(\tilde{y},\tilde{w})$. By Corollary 3.1, $\tilde{z}$ is also the unique minimum of

$\bar{\theta}_\rho(\cdot,\tilde{y},\tilde{w})$, the exact penalty function for $\bar{\mathcal{C}}(\tilde{y},\tilde{w})$. If $\tilde{z} \neq \tilde{y}$, then $\bar{\theta}_\rho(\tilde{z},\tilde{y},\tilde{w}) < \bar{\theta}_\rho(\tilde{y},\tilde{y},\tilde{w})$. It follows that for each y^* in $\partial_1\bar{\theta}_\rho(\tilde{y},\tilde{y},\tilde{w})$ we have

$$0 > \bar{\theta}_\rho(\tilde{z},\tilde{y},\tilde{w}) - \bar{\theta}_\rho(\tilde{y},\tilde{y},\tilde{w}) \geq \langle y^*,\tilde{z}-\tilde{y}\rangle = \langle y^*,d\rangle,$$

where $d = \tilde{z}-\tilde{y}$. Hence, by Lemma 4, we have

$$D_d\bar{\theta}_\rho(\tilde{y},\tilde{y},\tilde{w}) = \max\{\langle y^*,d\rangle \,|\, y^* \in \partial_1\bar{\theta}_\rho(\tilde{y},\tilde{y},\tilde{w})\} < 0 .$$

Since $\partial_1\bar{f}_k(\tilde{y},\tilde{y},\tilde{w}) = \partial f_k(\tilde{y})$ for $k = 0,1,\ldots,p$ and $\bar{f}_k(\tilde{y},\tilde{y},\tilde{w}) = f_k(\tilde{y})$ for $k = 1,2,\ldots,p$, it follows from Lemmas 3 and 4 and Theorem 4 that $D_d\theta_\rho(\tilde{y}) = D_d\bar{\theta}_\rho(\tilde{y},\tilde{y},\tilde{w}) < 0$, which proves the theorem. ⊗

The heart of our convergence theorem is the following slight generalization of Zangwill's convergence theorem [32].

<u>LEMMA 10</u>. Let Y be a nonempty and compact subset of R^m and let W be a compact subset of R^q. Let $\Gamma: Y \times W \to Y \times W$ be a point-to-set map. Suppose an algorithm generates the sequence $\{(x_i,w_i)\}$ according to the recursion $(x_{i+1},w_{i+1}) \in \Gamma(x_i,w_i)$, where (x_0,w_0) is given. Suppose that

1) there is a continuous function $\theta: Y \to R$ such that
   i) if x minimizes θ, then the algorithm stops at x
   ii) if x does not minimize θ, then whenever $(y,u) \in \Gamma(x,w)$ we have $\theta(y) < \theta(x)$
2) Γ is closed on $Y \times W$.

Then either the algorithm stops at some point (z,w) such that z minimizes θ, or some subsequence converges to some (z,w) such that z minimizes θ.

Proof. The proof does not differ significantly from that in [32], and will be omitted. ⊗

Though Zangwill's result guarantees only subsequential convergence, the following lemma provides a sufficient condition for the entire sequence $\{x_i\}$ to converge.

LEMMA 11. Let f be a convex function with the unique minimum z. If the sequence $\{x_i\}$ satisfies $\lim_{i \to \infty} f(x_i) = f(z)$, then $\lim_{i \to \infty} x_i = z$.

Proof. See Corollary 27.2.2, Rockafellar [26]. ⊗

LEMMA 12. Let $M_1 : X \to Y$ and $M_2 : Y \to Z$ be point-to-set maps. Let the composition map $M_2 M_1 : X \to Z$ be defined by

$$M_2 M_1(x) = \cup \{M_2(y) \mid y \in M_1(x)\} .$$

Suppose that M_1 is closed at x and M_2 is closed on $M_1(x)$. If Y is compact, then $M_2 M_1$ is closed at x.

Proof. See Corollary 4.2.1, Zangwill [32]. ⊗

LEMMA 13. Let $f : R^m \to R$ be a continuous function and let the point-to-set map $M : R^m \times R^m \to R^m$ be defined by

$$M(x,d) = \{y \mid f(y) = \min_{0 \le \alpha \le \beta} f(x + \alpha d)\} ,$$

where β is a fixed positive number. Then M is closed on $R^m \times R^m$.

Proof. See Lemma 5.1, Zangwill [32]. $\otimes$

LEMMA 14. Let $f_1, f_2, \ldots, f_p$ be convex functions such that $\{x | f_k(x) \leq 0,\ k = 1,2,\ldots,p\}$ is nonempty and bounded. Then for each real number α the level set $X_\alpha = \{x | \sum_{k=1}^{p} \max(0, f_k(x)) \leq \alpha\}$ is compact if it is nonempty.

Proof. See Lemma 3.4, Han [7].

We now define the algorithm. Let the positive numbers ρ and β, the nonempty and compact set $T \subset R^m$, and the compact set $W \subset R^q$ be given. Choose any (x_0, w_0) in $T \times W$. Consider the following idealized algorithm.

Algorithm $\mathcal{A}$: For $i = 0,1,2,\ldots$

1) solve $\mathcal{C}(x_i, w_i)$ to obtain a solution z_i; let $d_i = z_i - x_i$
2) find an α_i such that $\theta_\rho(x_i + \alpha_i d_i) = \min\{\theta_\rho(x_i + \alpha d_i) | 0 \leq \alpha \leq \beta\}$; let $x_{i+1} = x_i + \alpha_i d_i$
3) stop if $x_{i+1} = x_i$; otherwise, return to 1) with x_{i+1} replacing x_i and any w_{i+1} in W replacing w_i.

We may now prove the global convergence theorem.

THEOREM 6. Suppose that program $\mathcal{C}$ is superconsistent, that its objective function f_0 is bounded below, that its feasible region $\{x \mid f_k(x) \le 0,\ k = 1,2,\ldots,p\}$ is bounded, and that it has the unique solution z. Let $\bar{f}_0, \bar{f}_1, \ldots, \bar{f}_p$ be functions jointly continuous on $R^m \times R^m \times R^q$ such that for each fixed y, w, and $k = 0,1,\ldots,p$ the function $\bar{f}_k(\cdot,y,w)$ is convex and $\partial_1 \bar{f}_k(y,y,w) = \partial f_k(y)$, and such that for each x, y, w, and $k = 1,2,\ldots,p$ we have $\bar{f}_k(x,y,w) \le f_k(x)$ and $\bar{f}_k(y,y,w) = f_k(y)$. Suppose that program $\bar{\mathcal{C}}(y,w)$ has a unique solution whenever $(y,w) \in R^m \times W$. Then there is a positive number ρ_0 such that, whenever $\rho \ge \rho_0$, algorithm $\mathcal{A}$ either stops at the unique solution z or $\lim_{i\to\infty} x_i = z$.

Proof. By Corollary 3.2, there is a positive number ρ_3 such that z is also the unique minimum of θ_ρ whenever $\rho \ge \rho_3$. Let f_0 be bounded below on R^m by $-\sigma$, and let

$$Y = \{x \mid \sum_{k=1}^{p} \max(0,f_k(x)) \le \max_{x\in T} [\frac{1}{\rho_3} (f_0(x) + \sigma) + \sum_{k=1}^{p} \max(0,f_k(x))]\}.$$

Clearly, we have $T \subset Y$. Also, since $\sum_{k=1}^{p} \max(0,f_k(z)) = 0$, we have $z \in Y$. By Lemma 14, Y is compact.

Let $S = Y \times W$ if $W \ne \emptyset$, and let $S = Y$ if $W = \emptyset$. Applying Corollary 3.1 to the compact set S, we conclude that there is a positive number $\bar{\rho}_3$ such that the minimum set of $\bar{\mathcal{C}}_\rho(\cdot,s)$ and the solution set of $\bar{\mathcal{C}}(s)$ coincide whenever $\rho \ge \bar{\rho}_3$ and $s \in S$. By hypothesis, this common set is the singleton $\{z(s)\}$. Let $\rho_0 = \max(1, \rho_3, \bar{\rho}_3)$. We will show that ρ_0 is the desired constant.

Choose (x_0, w_0) in $T \times W$, choose $\rho \geq \rho_0$, and let $Y_0 = \{x | \theta_\rho(x) \leq \theta_\rho(x_0)\}$. Notice that $z \in Y_0$, since $\rho \geq \rho_3$. We will show inductively that, if Algorithm $\mathcal{A}$ generates the sequence $\{(x_i, w_i)\}$ (possibly a finite sequence), then $\{x_i\} \subset Y_0 \cap Y$. It is clearly true for x_0. Suppose $x_j \in Y_0 \cap Y$ for $j = 1,2,\ldots,i$, and let $z_i = z(x_i, w_i)$.

Suppose first that $z_i = x_i$. Then $d_i = 0$, $x_{i+1} = x_i$, and the algorithm stops. On the other hand, suppose that $z_i \neq x_i$. By Theorem 5, $d_i = z_i - x_i$ is a descent direction for θ_ρ at x_i. Therefore, the line search must generate an x_{i+1} such that $\theta_\rho(x_{i+1}) \leq \theta_\rho(x_i)$. From the induction hypothesis, we have $\theta_\rho(x_i) \leq \theta_\rho(x_0)$. Hence, $\theta_\rho(x_{i+1}) \leq \theta_\rho(x_i) \leq \theta_\rho(x_0)$, that is, $x_{i+1} \in Y_0$. It follows that

$$\sum_{k=1}^{p} \max(0, f_k(x_{i+1}))$$

$$\leq \frac{1}{\rho} (\theta_\rho(x_0) - f_0(x_{i+1})) = \frac{1}{\rho} (f_0(x_0) - f_0(x_{i+1})) + \sum_{k=1}^{p} \max(0, f_k(x_0))$$

$$\leq \frac{1}{\rho} (f_0(x_0) + \sigma) + \sum_{k=1}^{p} \max(0, f_k(x_0))$$

$$\leq \frac{1}{\rho_3} (f_0(x_0) + \sigma) + \sum_{k=1}^{p} \max(0, f_k(x_0))$$

(since $f_0(x_0) + \sigma \geq 0$ and $\rho \geq \max(1, \rho_3)$)

$$\leq \max_{x \in T}\Big[\frac{1}{\rho_3} (f_\rho(x) + \sigma) + \sum_{k=1}^{p} \max(0, f_k(x)\Big] \quad \text{(since } x_0 \in T) .$$

Therefore, $x_{i+1} \in Y_0 \cap Y$.

Let the map $D:S \to R^m \times R^m$ be defined by $D(s) = \{(x, z(s)-x)\}$, where $s = (x,w)$, and let the map $L:R^m \times R^m \to S$ be defined by

$$L(x,d) = \{(x+\alpha d, w) \mid \theta_\rho(x+\alpha d) = \min_{0 \leq \alpha \leq \beta} \theta_\rho(x+\alpha d) \text{ and } w \in W\}.$$

Lastly, let the composition map $\Gamma:S \to S$ be defined by $\Gamma = LD$. Clearly, if $s_i \in S$, then algorithm $\mathcal{A}$ generates the point s_{i+1} only if $s_{i+1} \in \Gamma(s_i)$.

We will verify that the hypotheses of Zangwill's Convergence Theorem are satisfied for the point-to-set map Γ and the continuous function θ_ρ. Actually, we have already shown that $\Gamma:S \to S$ and that S is compact.

Suppose that, for some i, the point x_i minimizes θ_ρ. Since $\rho \geq \rho_3$, it follows by Corollary 3.2 that x_i also solves $\mathcal{C}$. By Lemma 6, there is a vector u of Lagrange multipliers such that (x_i,u) is a K.K.T. pair for $\mathcal{C}$. Since $\bar{f}_k(x_i,x_i,w_i) = f_k(x_i)$ for $k = 1,2,\ldots,p$ and $\partial_1\bar{f}_k(x_i,x_i,w_i) = \partial f_k(x_i)$ for $k = 0,1,\ldots,p$, it follows that (x_i,u) is also a K.K.T. pair for $\bar{\mathcal{C}}(x_i,w_i)$ for any w_i in W. Therefore $x_i = z_i$, the unique solution of $\bar{\mathcal{C}}(x_i,w_i)$. Hence $d_i = z_i - x_i = 0$, and the algorithm stops.

On the other hand, suppose that x_i does not minimize θ_ρ. Reasoning as above, it follows that x_i does not solve $\mathcal{C}$, and hence x_i does not solve $\bar{\mathcal{C}}(x_i,w_i)$ for any w_i in W. Therefore, the unique solution z_i of $\bar{\mathcal{C}}(x_i,w_i)$ must satisfy $z_i \neq x_i$. Since we assumed that an exact line search is executed over a nonempty interval, it follows from Theorem 5 that $\theta_\rho(x) < \theta_\rho(x_i)$ whenever $(x,w) \in \Gamma(x_i,w_i)$.

Let $\Omega: S \to R^m$ be defined by $\Omega(s) = \{z(s)\}$. Then Ω is closed on S and uniformly bounded near S, by Lemma 1. Therefore, by Lemma 2, for some finite number b we have $\sup\{\|z(s)\| \mid s \in S\} \leq b$. Let $B = \{y \in R^m \mid \|y\| \leq b\}$. Then, for each $s = (x,w)$ in S, the pair $(s, z(s)-x)$ is contained in the compact set $S \times (B-Y)$. Therefore, by Lemmas 1, 12, and 13, the map Γ is closed on S. Thus the hypotheses of Lemma 10 are satisfied; we conclude that either the algorithm stops at z or some subsequence $\{x_j\}$ converges to z. Since the entire sequence $\{\theta_\rho(x_i)\}$ is monotone decreasing, by Lemma 9 we have

$$\lim_{i \to \infty} \theta_\rho(x_i) = \lim_{j \to \infty} \theta_\rho(x_j) = \theta_\rho(z) = f_0(z) .$$

It follows from Lemma 11 that $\lim_{i \to \infty} x_i = z$, which proves the theorem. ⊗

6. Concluding Remarks

It is clear from the proof of Theorem 6 that ρ_0 depends on T and W but not on β. Although the theorem holds for each positive number β, in practice β should be chosen suitably large in the hopes of insuring that the line search terminates because the minimum is reached, and not because the upper bound β is encountered. Such a choice could only speed the overall convergence.

The requirement that f_0 be bounded below can always be met by replacing f_0 with $\exp(f_0)$, which is bounded below by zero. If $\mathcal{C}$ has a unique solution, then $\mathcal{C}$ will also have a unique solution when $\exp(f_0)$ replaces f_0.

If $\mathcal{C}$ has the unique solution z, then in theory we can always insure that the feasible region is bounded by imposing the single additional constraint $\langle x,x \rangle \leq c$, where $c > \langle z,z \rangle$. In practice, a very large value of c should be used. Alternatively, we could bound the feasible region with linear constraints.

The most restrictive hypothesis is the requirement that each $\bar{\mathcal{C}}(y,w)$ possess a unique solution. For quadratic subproblems, this is accomplished by using a positive definite matrix. Notice that global convergence is assured even if one fixed positive definite matrix is used for each quadratic subproblem. However, the local properties of the algorithm will then suffer.

To study the local behavior of a recursive substitution scheme, it is usual to make strong assumptions, including the requirement that each f_k and $\bar{f}_k$ be twice differentiable, and that a good estimate of a Karush-Kuhn-Tucker pair (z,u) be available. Under such conditions, analysis of a recursive substitution scheme utilizing quadratic subproblems [5,6,22], or arbitrary approximating subproblems in a one-point scheme [25], has shown that near (z,u) the line search can be omitted, and the resulting pure recursive substitution scheme generates a sequence $\{(x_i,v_i)\}$ that converges to (z,u). Moreover, a linear, superlinear, or quadratic convergence rate is possible, depending on second order conditions. Notice that the multiplier estimates v_i play a crucial role in the local analysis, yet are not explicitly considered in our global convergence theorem.

We hope that our global convergence result, motivated by the need to validate an algorithm for geometric programming [28], will inspire additional work in non-quadratic subproblems. In addition, the results in Sections 2 and 3 suggest a way to solve convex programs with nondifferentiable constraints. Namely, minimize the exact penalty function associated with the program, using any available algorithm for minimizing a nondifferentiable convex function [1,12].

Acknowledgment

The author benefited greatly from discussions with Margaret Wright, who also made helpful suggestions on a preliminary draft of this paper.

REFERENCES

[1] M. L. Balinski and P. Wolfe, editors, Mathematical Programming Study 3: Nondifferentiable Optimization, North Holland, Amsterdam, 1975.

[2] J. Danskin, The Theory of Max-Min, Springer-Verlag, New York, 1967.

[3] R. J. Duffin, "Linearizing Geometric Programs," SIAM Review, 12, 1970, pp. 211-227.

[4] A. F. Fiacco and G. P. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, Wiley, New York, 1968.

[5] U. M. Garcia Palomares and O. L. Mangasarian, "Superlinearly Convergent Quasi-Newton Algorithms for Nonlinearly Constrained Optimization Problems," Mathematical Programming, 11, 1976, pp. 1-13.

[6] S. P. Han, "Superlinearly Convergent Variable Metric Algorithms for General Nonlinear Programming Problems," Mathematical Programming, 11, 1976, pp. 263-282.

[7] S. P. Han, "A Globally Convergent Method for Nonlinear Programming," J. of Opt. Th. and Appl., 22, 1977, pp. 297-309.

[8] S. P. Han and O. L. Mangasarian, "Exact Penalty Functions in Nonlinear Programming," M.R.C. Technical Summary Report 1897, Math. Research Center, Madison, Wisconsin, 1978.

[9] W. Hogan, "Point-to-Set Maps in Mathematical Programming," SIAM Review, 15, 1973, pp. 591-603.

[10] W. Hogan, "Directional Derivatives for Extremal-Value Functions with Applications to the Completely Convex Case," Operations Research, 21, 1973, pp. 188-209.

[11] P. Huard, "Optimization Algorithms and Point-to-Set Maps," Mathematical Programming, 8, 1975, pp. 308-331.

[12] C. Lemarechal, "Minimization of Nondifferentiable Functions with Constraints," in Proc. Twelfth Annual Allerton Conferences on Circuit and System Theory, 1974, pp. 16-24.

[13] E. S. Levitin and B. T. Polyak, "Constrained Minimization Methods," U.S.S.R. Computational Mathematics and Mathematical Physics, 6, 1966, pp. 1-50.

[14] D. G. Luenberger, "Control Problems with Kinks," IEEE Transactions on Automatic Control, Vol. 15-AC, No. 5, 1970, pp. 570-575.

[15] D. G. Luenberger, Introduction to Linear and Nonlinear Programming, Addison-Wesley, Reading, Mass., 1973.

[16] R. R. Meyer, "The Validity of a Family of Optimization Methods," SIAM J. Control, 8, 1970, pp. 41-54.

[17] R. R. Meyer, "Sufficient Conditions for the Convergence of Monotonic Mathematical Programming Algorithms," J. of Computer and System Sciences, 12, 1976, pp. 108-121.

[18] R. R. Meyer, "A Comparison of the Forcing Function and Point-to-Set Mapping Approaches to Convergence Analysis," SIAM J. Control and Optimization, 15, 1977, pp. 699-715.

[19] W. Murray, Constrained Optimization, Report MA79, National Physical Laboratory, 1969.

[20] W. Murray, "An Algorithm for Constrained Minimization," in Optimization (edited by R. Fletcher), Academic Press, New York, 1969, pp. 247-258.

[21] T. Pietrzykowski, "An Exact Potential Method for Constrained Maxima," SIAM J. Numer. Anal., 6, 1969, pp. 299-304.

[22] M. J. D. Powell, "The Convergence of Variable Metric Methods for Nonlinearly Constrained Optimization Calculations," in Nonlinear Programming 3 (edited by O. L. Mangasarian, R. R. Meyer, and S. M. Robinson), Academic Press, New York, 1978, pp. 27-63.

[23] B. N. Pshenichnyi, "Convex Programming in a Normalized Space," Kibernetika, 1, 1965, pp. 46-57.

[24] S. M. Robinson, "A Quadratically Convergent Algorithm for General Nonlinear Programming Problems," Mathematical Programming, 3, 1972, pp. 145-156.

[25] S. M. Robinson, "Perturbed Kuhn-Tucker Points and Rates of Convergence for a Class of Nonlinear-Programming Algorithms," Mathematical Programming, 7, 1974, pp. 1-16.

[26] R. T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, New Jersey, 1970.

[27] J. B. Rosen and J. Kreuser, "A Gradient Projection Algorithm for Nonlinear Constraints," in Numerical Methods for Nonlinear Optimization (edited by F. A. Lootsma), Academic Press, New York, 1972, pp. 297-300.

[28] E. Rosenberg, "A Globally Convergent Condensation Algorithm for Geometric Programming," in preparation.

[29] R. B. Wilson, "A Simplicial Algorithm for Concave Programming," Dissertation, Graduate School of Business Administration, Harvard University, Cambridge, Mass., 1963.

[30] M. A. H. Wright, Numerical Methods for Nonlinearly Constrained Optimization, Report No. 193, Stanford Linear Acceleration Center, 1976.

[31] W. Zangwill, "Non-Linear Programming via Penalty Functions," Management Science, 13, 1967, pp. 344-358.

[32] W. Zangwill, Nonlinear Programming: A Unified Approach, Prentice-Hall, Englewood Cliffs, New Jersey, 1969.

[33] G. Zoutendijk, Methods of Feasible Directions, Elsevier, Amsterdam, 1960.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
14 SOL-79-1

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
6 GLOBALLY CONVERGENT ALGORITHMS FOR CONVEX PROGRAMMING.

5. TYPE OF REPORT & PERIOD COVERED
9 Technical Rep[redacted]t.

6. PERFORMING ORG. REPORT NUMBER
SOL 79-1

7. AUTHOR(s)
10 Eric Rosenberg

8. CONTRACT OR GRANT NUMBER(s)
15 N00014-75-C-0267, NSF-MCS76-81259

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Department of Operations Research -- SOL
Stanford University
Stanford, CA 94305

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
NR-047-143

11. CONTROLLING OFFICE NAME AND ADDRESS
Operations Research Program -- ONR
Department of the Navy
800 N. Quincy St., Arlington, VA 22217

12. REPORT DATE
11 Febr[redacted]979

13. NUMBER OF PAGES
38

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
12 42p

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Convergent Programming
Non-differentiable Optimization
Global Convergence

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

SEE ATTACHED

408 765 Gm

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SOL 79-1 by Eric Rosenberg
GLOBALLY CONVERGENT ALGORITHMS FOR CONVEX PROGRAMMING

We consider solving a (minimization) convex program by sequentially solving a (minimization) convex approximating subproblem and then executing a line search. Each subproblem is constructed from the current estimate of a solution of the given problem, possibly together with other information. Under mild conditions, solving the current subproblem generates a descent direction for an exact penalty function. Minimizing the exact penalty function along the current descent direction provides a new estimate of a solution, and a new subproblem is formed. For any arbitrary starting estimate, this scheme generates a sequence of estimates that converges to a solution of the given problem. Moreover, it is not necessary to require that the functions defining the given problem and each subproblem be differentiable.